END

FILMED

DTIC

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ARO 18623.16-MA	2. GOVT ACCESSION NO. N/A	3. RECIPIENT'S CATALOG NUMBER N/A
4. TITLE (and Subtitle) Construction of Efficient Algorithms for the Estimation of Multivariate Probability Densities		5. TYPE OF REPORT & PERIOD COVERED Final 11/1/81-10/31/84
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) James R. Thompson		8. CONTRACT OR GRANT NUMBER(s) DAAG-29-82-K-0014
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Mathematical Sciences Rice University Houston, Texas 77251-1892		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office Post Office Box 12211 Research Triangle Park, NC 27709		12. REPORT DATE 1/4/85
		13. NUMBER OF PAGES 6
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

NA

18. SUPPLEMENTARY NOTES

The view, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Average shifted histogram, nonparametric probability density estimation, simulated quasi maximum likelihood, modes of high dimensional densities, multicolour time lapsed contour display, EM algorithm, Gaussian mixtures, data based simulation, behind armour data, birth and death process, metastatic progression, stochastic processes, remote sensing, satellite overflights.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

State of the art software has been created for the nonparametric estimation and contour display of densities of dimensions through five. A nonparametric mode finding algorithm has been developed for densities of dimension up to one hundred. An algorithm for data based simulation from high dimensional distributions has been created and employed on behind armour data. In the related area of parameter estimation of stochastic processes, simulation based algorithms have been developed which enable the user to proceed directly from the axioms of the process plus data to the estimation of parameters. Additional research in progress
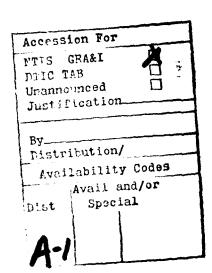
DD 1 JAN 73 1473

85 03 07 044

CONSTRUCTION OF EFFICIENT ALGORITHMS FOR THE ESTIMATION

OF MULTIVARIATE PROBABILITY DENSITIES

JAMES R. THOMPSON

U.S. ARMY RESEARCH OFFICE

DAAG-29-82-K-0014

RICE UNIVERSITY

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED

Statement of Problem Studied

Our task has been to develop algorithms, and, to whatever extent possible, portable software for the data based nonparametric estimation of densities of high dimensionality. We hoped to employ graphical and interactive techniques whenever possible. In cases where the dimensionality was very high, we hoped to develop an algorithm for the finding of modes which might then be used as origins for further investigations. In addition, we hoped to examine related problem areas, such as parameter estimation for applied stochastic process models. It was hoped that interaction with the Ballistics Research Laboratory at Aberdeen would be fruitful. Also, it was hoped that the Texas Medical Center and NASA would provide useful local data sources on which to test our algorithms.

Summary of Most Important Results

The main thrust of our work has been in the nonparametric estima-
tion and representation of densities of dimensions of three and greater.
An algorithm for multicolour, time lapsed density contour display has
been developed by Scott and is now portable. A 16mm film has been made
of the application of this technique to estimate and display densities
of dimensions three, four and five. This film has been shown at the
Aberdeen Proving Ground, at the 1983 DOD Conference at the USAF Academy,
at the ONR 1983 Luray Conference, in addition to some dozen nonDOD
conferences. The principal data sets used in explication of the Scott
algorithm are ground observations from satellite overflights.

The nonparametric estimation and representation of densities of
very high dimensionality (up to one hundred) is a problem addressed in
the doctoral dissertation of Steven Boswell, presently of Lincoln
Laboratories. Boswell's algorithm seeks out the modes of the density
which generated a data set. These modes are then to be used as origins
for further analysis.

The dissertation of Hathaway uses the EM algorithm to address the
problem of estimating a density as a mixture of Gaussian densities.
Many of the instablity problems generally encountered with mixture tech-
niques are eliminated.

The problem of using a data base as a vehicle for simulation has
been addressed by Taylor and Thompson using behind armour data from the
Aberdeen Proving Ground. Their procedure gives the same effects as those
which would have been obtained if the underlying density had first been
estimated nonparametrically and then used to build a random number gen-
erator. However, by eliminating the density estimation step, the
Taylor-Thompson algorithm provides enormous savings both in computer
running times, and, more importantly, in the ease of use.

The development of simulation techniques for parameter estimation
in complex stochastic process situations has been a major activity for
Thompson and his colleagues at the Texas Medical Center, Atkinson, Bar-
toszynski and Brown. Not only are these techniques more stable and
quicker than those based on approximation theoretic procedures, but,
since they require only the axioms of the process, rather than some
"closed form solution," they are much easier to use. A problem which
had required a man year to code using standard techniques was coded in
one week using the simulation based strategy. The development of this
algorithm enables the use of stochastic models in real world situations
which earlier had defied such applied modeling.

Publications

Atkinson,E.N., Bartoszynski,R., Brown,B.W., Thompson,J.R.(1983) "Simulation Techniques for Parameter Estimation in Tumor Related Stochastic Processes." Proceedings of the 1983 Summer Computer Simulation Conference, pp. 754-757.

Atkinson,E.N., Bartoszynski,R., Brown,B.W., Thompson,J.R. (1983) "Maximum Likelihood Techniques in Cancer Modeling." Proceedings of the 1983 Symposium of the International Statistical Institute, pp. 494-497.

Atkinson,E.N., Bartoszynski,R., Brown,B.W., Thompson,J.R.(1984) "On estimating the growth function of tumors." Mathematical Biosciences, v.67, pp.145-166

Atkinson,E.N., Bartoszynski,R., Brown,B.W., Thompson,J.R.(1984) "Estimation of tumor growth rate from distributrion of size at detection." Journal of the National Cancer Institute, v.72, pp. 31-38.

Bartoszynski, Brown and Thompson.(1982) "Metastatic and Systemic Factors in Neoplastic Progression," in Probability Models and Cancer, LeCam and Neyman, eds. North-Holland, pp. 253-263.

Boswell, Steven B. (1983) Nonparametric Mode Estimation for Higher Dimensional Densities. Ph.D. Dissertation (Thompson,J.R., thesis director).

Boswell, Steven B. (1984) "Nonparametric Estimation of the Modes of High Dimensional Densities," to appear in Proceedings of the 1984 Conference on the Interface between Computer Science and Statistics.

Chiu, Shean-Tsong (1984) "Statistical estimation of the parameters of a moving source from array data," submitted.

Hathaway, Richard (1982) The EM Algorithm for the Maximum Likelihood Estimation of Gaussian Mixtures. Ph.D. Dissertation (Thompson,J.R. and Dennis, J.E. thesis directors).

Scott,D.W. and Thompson,J.R. (1983) "Probability density estimation in higher dimensions." In the Proceedings of the Fifteenth Symposium on the Interface of Computer Science and Statistics (James E. Gentle, ed.), pp.173-179.

Scott,D.W. (1983) "Multivariate density estimation and remote sensing," in Mathematical Pattern Recognition and Image Analysis, pp. 77-92.

Scott, D.W. (1984) "Multivariate Density Function Representation," in _Proceedings of the Fifth Annual National Computer Graphics Association Conference_, Volume II, 794-801, May, 1984.

Scott, D.W. "Frequency Polygons: Theory and Application." To appear in _J. American Statistical Association_.

Scott, D.W. "Data Analysis in 3 and 4 Dimensions With Nonparametric Density Estimation," in _Statistical Image Processing_. E.J. Wegman and D. DePriest, Eds., Marcel Dekker, New York, in press.

Scott, D.W. "Averaged Shifted Histograms: Effective Nonparametric Estimators in Several Dimensions," tentatively accepted, _The Annals of Statistics_, November, 1984.

Scott, D.W. and Jee, R. "Nonparametric Analysis of Minnesota Spruce and Aspen Tree Data and Landsat Data," to appear in _Proceedings of the Second Symposium on Mathematical Pattern Recognition and Image Analysis_, June, 1984.

Scott, D.W. and S.J. Sheather. "Kernel Density Estimation with Binned Data," submitted.

Scott, D.W. (1983) "Nonparametric Probability Density Estimation for Data Analysis in Several Dimensions," in _Proceedings of the Twenty-Eighth Conference on the Design of Experiments in Army Research Development and Testing_, pp. 387-397.

Scott, D.W. and Terrell, G.R. "Oversmoothed Nonparametric Density Estimates," to appear in _J. American Statistical Association_.

Scott, D.W. "Optimal Meshes for Histograms Using Variable-Width Bins," submitted.

Taylor, M.S. and Thompson, J. R. "Data Based Random Number Generation for a Multivariate Distribution via Stochastic Interpolation," submitted.

Thompson, J.R. (1984) "Deterministic versus stochastic modeling in neoplasia," in _Proceedings of the 1984 Computer Simulation Conference_, pp. 822-825.

Scientific Personnel Supported and Ph.D. Degrees Awarded

Bartoszynski, Robert: Visiting Professor of Mathematical Sciences

Boswell, Steve: Graduate Student (presently Research Associate Lincoln Labs), Ph.D. awarded 1983.

Chiu, Shean-Tsong: Assistant Professor of Mathematical Sciences

Hathaway, Richard: Graduate Student (presently Assistant Professor Univ. of S.C.) Ph.D. awarded 1982.

Jee, Rodney: Graduate Student (hired by JPL), Ph.D. to be completed in 1985

Scott, David W.: Associate Professor of Mathematical Sciences

Stevens, Joyce: Graduate Student.

Tapia, Richard A.: Professor of Mathematical Sciences

Terrell, George R. : Visiting Assistant Professor of Mathematical Sciences

Thompson, James R.: Professor of Mathematical Sciences

END

FILMED

4-85

DTIC